

SOME PROBLEMS ARISING FROM THE PRELIMINARY MANIPULATION OF PLANT ECOLOGICAL DATA FOR SUBSEQUENT NUMERICAL ANALYSIS

B. H. WALKER

(Division of Biological Sciences, University of Rhodesia)

ABSTRACT

The increasing use of multivariate methods in analysis of ecological data demands that more attention be paid to the validity of results being obtained. An area of weakness in the application of these techniques lies in the form that the data should take before they are analyzed. Initial manipulation of the data, or the lack thereof, strongly influences the outcome of the analysis. This paper is for the interest of non-mathematical ecologists and attempts to describe some of the problems associated with importance values, the presence of zero-values, standardization of data, the use of relative values, and distortions due to non-linearity of the data. It is concluded that there is seldom a best method of transformation and analysis, and that for any particular analytical procedure more than one analysis should be made on different transformations of the data. Although producing apparently different answers, each may be valid, emphasizing different aspects of the data.

UITTREKSEL

PROBLEME AS GEVOLG VAN DIE VOORAFGAANDE MANIPULASIE VAN PLANTEKOLOGIESE DATA VIR NUMERIESE ANALISE

Die toenemende gebruik van veelvuldige veranderlike metodes by die ontleding van ekologiese gegewens vereis dat meer aandag bestee moet word aan die geldigheid van uitslae wat verwerf word. 'n Swakheid by die toepassing van hierdie tegnieke lê by die vorm wat die gegewens behoort aan te neem voordat hulle ontleed word. Aanvanklike bewerking van die gegewens, of die gebrek daaraan beoefen 'n sterk invloed oor die uitslag van die ontleding uit. Hierdie verhandeling word gerig aan nie-wiskundige ekoloë en poog om sommige van die probleme verbonde aan gewigswaardes, die aanwesigheid van nulwaardes, standaardisering van gegewens, die gebruik van relatiewe waardes en distorsie weens nie-lineêre gegewens, te beskryf. Dit word afgelei dat daar selde 'n besondere metode van omskepping en ontleding is. Vir enige bepaalde analitiese werkswyse behoort meer as een ontleding op verskillende omskeppings van gegewens gemaak te word. Alhoewel dit klaarblyklik verskillende antwoorde afgee, mag elk geldig wees deurdat verskillende aspekte van die gegewens beklemtoon word.

INTRODUCTION

The rapid increase in development and application of numerical methods in plant ecology carries with it considerable promise for improving the understanding of vegetation/environment relationships. However, it also gives rise to a number of problems, associated with the basic nature of ecological data, that can lead to invalid results. It is unfortunate that in many instances research workers who use multivariate techniques have insufficient background to properly evaluate the array of methods with which they are confronted, and

since it is neither desirable nor possible to offer a cut-and-dried recipe for ecological research, there is a need for general, non-mathematical accounts of common problems that arise in the use of numerical methods. The great danger is that a lack of appreciation of basic assumptions about the nature of the data can lead to spurious patterns emerging from the analysis, and being accepted by a researcher whose wide range of activities prevents him from devoting the time required to fully comprehend the output of a complex analysis.

This paper is not aimed at statistical ecologists, and is not intended to be a critical account of multivariate analyses. Its objectives are to demonstrate some of the common problems arising from the preparation of data for subsequent numerical analysis. It is hoped that it may help to clarify the situation for those who wish only to use the methods for what they really are—ecological research tools. The rationale of particular methods is omitted (except where pertinent to alternatives in data transformation) and attention is limited to the suitability of various forms of data. For those who wish to delve deeper, the mathematical bases for transformation in general statistical procedures are to be found in Patil *et al.* (1971), and the papers by Williams (1971) and Austin and Greig-Smith (1968) should also be consulted. For the field ecologist, Greig-Smith (1971a) and Dagnelie (1971) present a non-mathematical account of what some of the procedures can and cannot do and cover the factors involved in the choice of technique for any particular problem. The review by Goodall (1970) summarizes the present status of statistical plant ecology.

One of the main weaknesses in the application of multivariate techniques to plant ecological data seems to be the loose definition of what the raw species values are supposed to represent, and the often arbitrary manipulation of these data (standardization, weighting, etc.) as a preliminary step in their analysis. Apart from the special case of presence and absence data the final scores given to each species in each of the sites (stands) will clearly play a major role in determining the nature of the results obtained. Certain alternatives within the chosen analytical procedure can produce marked changes in results, for example in cluster analysis by varying the choice of fusion strategy (Grigal and Goldstein, 1972). However, problems relating to choice of analytical procedure are outside the scope of this paper. The following discussion concerns problems relating to the form of the data that are likely to cause distortions in the results, whatever the method of analysis.

THE PROBLEM OF IMPORTANCE VALUES

Whichever method is used to analyse the data matrix the assumption is made that the values therein accurately reflect the ecological importance of the species. The first question to answer, therefore, is which measure or combination of measures of the vegetation truly represents the importance of the species?

The answer lies in the aims of the study, and until some definitive work allows for an objective selection, the choice of measure must remain a pragmatic one (Goff and Cottom, 1967; Moore *et al.*, 1970) based on the requirement that it will supply the information about the vegetation that is needed to meet the aims of the study.

More than one measure may be needed to adequately represent the species, although the use of such composite measures has been opposed (Langford and Buell, 1969) on the basis that they can conceal significant interstand differences, for example where two species have inverse relationships with respect to density and basal area. But in such a case several questions arise. Are 20 bigger trees equal in importance to 40 smaller trees? And if not, is it density or basal area which reflects more closely the ecological significance of the species? Austin and Greig-Smith (1968) have demonstrated how the first ordination axis from an analysis of tropical rain forest data changes according to whether density or basal area is used to represent the species. Without recourse to an objective selection of one or the other measures it is generally better, for example, to weight in some way the density value by the size of individuals, than to use density alone while knowing full well that the individuals are of different sizes. Perhaps an appropriate general importance value for most phytosociological studies would be biomass or some kind of functional value, such as the amount of CO₂ fixed per day by each species, (Whittaker, 1965). But for studies where succession or stage of development is of particular interest (reflected in size of individuals?) it may not be so useful. Being restricted to the characters that can be measured in practice, it may be that certain species are best represented by density, whereas for others (e.g. a rhizomatous grass species) density may have little meaning. Since both kinds of species have to be included in the analysis, they may require different measures, and their comparison is therefore complicated even further. Only if the data are to be subsequently standardized will the direct use of different measures for different species be valid. If the species are to retain their differences in variance, the units used for the two measures must be scaled to the same order of magnitude. The relative pros and cons of standardization are discussed later.

Lambert and Dale (1964) have recommended the use of presence and absence data only, thereby eliminating problems of importance values. In certain types of broad studies where species variation is high and a large number of stands is required, this may be the best approach. In such cases the use of quantitative estimates may not add significantly to the ecological interpretation obtained by analysis of the presence and absence data. However, the possibility of distortion through a chance presence or absence will always remain a danger with this approach, and the objectivity of the method may have to be sacrificed for common sense, to correct blatant anomalies (Moore *et al.*, 1970; Walker

and Coupland, 1970). Where the species of particular interest occur in most of the stands, the information contained in their differences of abundance in the stands cannot be ignored. Two species may occur in all stands and yet exhibit a negative correlation as a result of opposite trends in abundance.

THE PRESENCE OF ZERO VALUES

The allocation of a zero to those species in a study that do not occur in a particular stand is to imply that these species have an equal probability (zero) of occurring in that stand. Since this is not true, to ignore the fact is ecologically more wrong (albeit more objective) than to substitute other values, generated from the remainder of the data, that indicate the degree of absence of these species (Swan, 1970).

If the raw data matrix contains few zero values the substituted values will be well defined. They are determined by averaging the association of absent species with those that are present in the stand, and with few zero values the new values are averages of many measured associations. However, since only a small percentage of the values are changed the effect of the transformation is not very great. As the number of zero's increases so the need for, and effect of, transformation increases, but the reliability of the substituted values de-

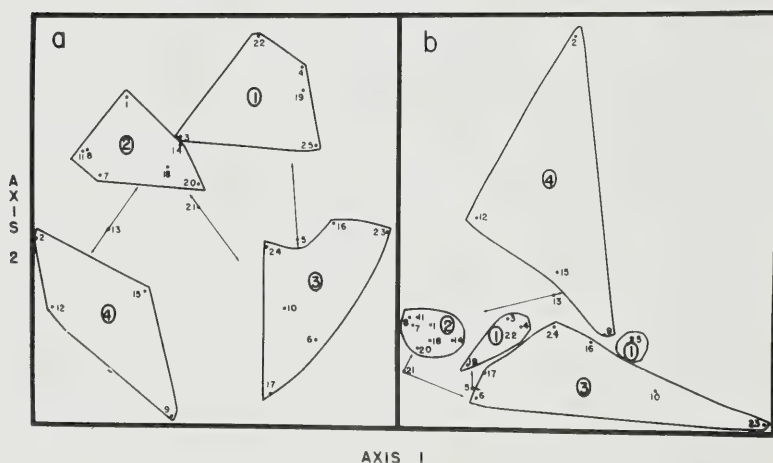


FIG. 1.

First two axes of an ordination by principal component analysis of 25 stands of herbaceous vegetation from southwestern Rhodesia. (a) Following zero-transformation, (b) using the raw data matrix. The data consist of the frequency distribution of 73 species in each stand. See text for explanation.

creases. Figures 1 and 2 demonstrate the effects of zero-transformation on an ordination by principal component analysis, and on a cluster analysis of a set of data from the Kalahari sand deposits in south-western Rhodesia. The data consist of frequency measures of 73 common herbaceous species in 25 stands, using 75 quadrats per stand. They are used only to demonstrate the effects of zero-transformation. The number of stands is insufficient for a full interpretation of the ecology of the region. The ordination procedure is that described by Walker and Wehrhahn (1971). The cluster analysis is an agglomerative, weighted pair group method based on a matrix of inter-stand correlation coefficients.

The four groups of stands depicted in Fig. 1 are as follows: (1) developed woodlands, relatively undisturbed, on deep, infertile sands; (2) scrubby (disturbed?) areas on the same site-types as 1; (3) soils with higher fertility, higher silt contents and often with a restricted rooting zone; (4) sites with a high water table, mostly treeless. Stands 13, 21 and 5 are intermediate with respect to site type, as indicated by the arrows. The first component represents a gradient from treeless stands through to the herbaceous layer that develops under woodland. The second is from deep sand sites (stands 22, 4) to those restricted in depth by either a hard layer that forms a barrier to roots (stands 6 and 17), or by a high water table (all of group 4). The pattern is much clearer in Fig. 1a, following zero-transformation. Fig 1b is apparently evincing the kind of distorted gradients described by Swan (1970).

With respect to the cluster analyses (Fig. 2), levels of correlation are higher using raw data, but zero-transformation has given more ecologically informative results. It has distinguished groups 3 and 4 (as described in the ordination analysis) from the combined groups 1 and 2, with stand 4 being apparently misplaced. It has therefore separated out the three different site types, but has left together the scrubby and developed forms of vegetation on deep, infertile sands. Cluster analysis of the raw data (Fig. 2b) has distinguished between the latter two groups but has failed to separate out the different site types, although group 4 can be recognised to some extent. The two analyses have emphasized different aspects of the vegetation, and each is valid. The use of zero-transformation results in a data set which is more continuous and therefore produces less sharply defined boundaries between groups, and it increases the emphasis in differences or similarities that exist in species with limited abundance. The groups produced by analysis of the raw data reflect differences and similarities in only the dominant and more abundant species. The effects of zero-transformation in cluster analysis will depend on the clustering strategy, and will be less marked if the technique disregards matched zero's.

In contrast to the above, cluster analyses of 44 stands of woody vegetation in the Harvard Forest, Massachusetts (Figs. 3a and 3b), showed the raw data

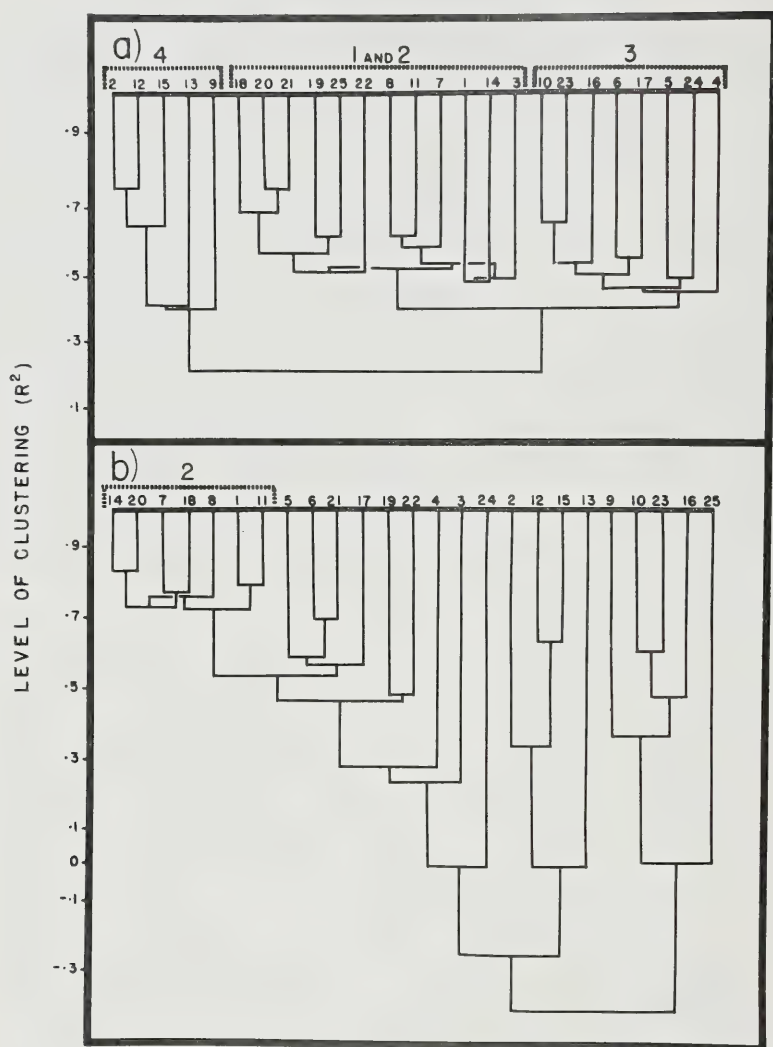


FIG. 2.

Cluster analysis of the same data used in Fig. 1. (a) Following zero-transformation, (b) using the raw data matrix. An agglomerative, weighted pair-group method is used, based on a matrix of inter-stand correlation coefficients.

to be more informative than zero-transformed data. The former separated out six community-types—namely hemlock (1), red maple swale (2), mixed red oak (3), hickory, (4), white pine (5) and black spruce (6), while analysis of the zero-transformed data led to groups 2, 3, and 4 being combined, and more stands being unplaced.

One possible drawback of zero transformation that has emerged from the above studies is that it tends to increase the variance of species with a high proportion of zero's, such that these species may contribute more to the total variance than the more abundant species.

STANDARDIZATION

Standardization of a raw data matrix is performed in order to make the variables comparable. For example, if two variables in the data are pH (measured in units of 1–14) and soil depth (varying perhaps from 20 to 500 cm) they cannot be compared in their original units in the same analysis. If, as in most analyses, the procedure involves successive reductions of variance by one means or another, the variables with the highest measures of variance will in general dominate the results. To make pH and soil depth comparable they should be expressed in the same units. There are various ways in which this can be done, and the most comprehensive account of the methods as they apply to ecological data is given by Austin and Greig-Smith (1968), who describe eight different techniques, including one which involves a zero-transformation.

The most common method is to divide the value of each variable in a stand by the standard deviation of that variable in the study as a whole. It is a conventional procedure adopted in most computer package programmes and it is implicit in any component analysis which is based on a species correlation coefficient matrix, since the correlation coefficient is a standardized measure of covariance. But with ecological data it can distort the outcome of the results, because the influence of all variables is made equal, and the contribution of the lesser species to the total variance is therefore increased. These species generally occur in only a few stands in small amounts and, while it may be that some of them are valuable indicators, they are mostly undersampled variables with little information.

If the data contain one or a few species with particularly high variances, standardization can have a significant adverse effect because the influence of these important species may be so diminished that the gradients they represent are merged with others which have little ecological meaning. But if most species are well represented, standardization will not have drastic, adverse effects. The Rhodesian data discussed earlier fall into this category, and a principal component analysis following standardization of the data that gave rise to Fig. 1a produced an almost identical configuration of the first two axes. Austin and



FIG. 3.

Cluster analysis of 44 stands of woody vegetation in the Harvard Forest, Massachusetts. (a) Using the raw data matrix, (b) following zero-transformation. The data consist of cover/abundance values for 23 species in each stand. The method of clustering is the same as for Figure 2.

Greig-Smith (1968) have concluded that, in general, standardization is desirable in species to avoid the abundance gradient which occurs with unstandardized data, and that the particular form of standardization only becomes important when there are numerous rare species.

If the importance value given to a species is its frequency of occurrence in a number of quadrats placed in a stand, then the resulting data set has values which vary between 0 and 100. The range of most species is comparable, and, subject to the drawbacks of unstandardized data just described, the data can be analyzed in their absolute form. But in some cases, for example where weighted importance values are created by combining estimates of basal area and height, the values may vary from 1 to over 1 000, and in this case a species whose maximum value is less than a hundred will have little influence on the analysis, even though it may occur in a high percentage of stands. In such a case the first, and sometimes second, gradient that appears in a component analysis is merely a reflection of the relative abundances of the species, and can be considered as no more than an abundance gradient. In discriminant function analysis, the opposite effect is obtained with respect to the values of the latent vectors (Hope, 1968). Table 1 gives the percentage contribution to the total variance of 35 woody species in 23 of the stands from Rhodesia (described earlier), using importance values of height in m $\times$ basal diameter in m. Intuitively, these values adequately represent the importances of the species, but in a multivariate analysis *Baikiaea plurijuga* and *Acacia giraffae* completely overshadow the influence of all the other species. Greig-Smith (1971b) obtained similar results from an ordination of data from tropical rain forests in the Solomon Islands. Two means of overcoming the problem, without resorting to standardization by equal variance, are as follows.

The data for each species can be adjusted to vary between 0 and 1, by dividing each value by the highest value for that species. Column (c) in Table 1 shows the effect of this on the same data used above. The situation is improved, but certain species such as *Commiphora africana* are now weighted in importance beyond their contribution to the total variance, even more so than in the case of standardization by equal variance.

The other approach is to transform the data logarithmically. Column (b) in Table 1 shows the effects of this on the same set of data as before, and in this particular instance it was, overall, the best solution, giving rise to the clearest interpretation of the results of a principal component analysis. It is of significance that although the raw data led to meaningless ordinations, the same interpretation was obtained by all three methods of standardization, but with differing degrees of clarity.

The use of a log. transform may not always have a desirable effect, especially if it follows zero-transformation. As described previously, the latter tends to

TABLE 1

Percentage contributions to the total variance of importance values of 35 woody species in 23 stands in the Wankie National Park, Rhodesia. (a) raw data, (b) log. of raw data, (c) data for each species transformed to the range of 0-1.

<i>Species</i>	(a)	(b)	(c)
<i>Acacia ataxacantha</i>	0,8	4,0	1,8
<i>A. fleckii</i>	0,3	1,0	1,4
<i>A. giraffae</i>	30,6	4,6	2,0
<i>A. tortilis</i>	3,1	6,9	1,9
<i>Allophyllus rhodesicum</i>	0,1	3,8	3,8
<i>Baikiaea plurijuga</i>	35,5	6,0	3,6
<i>Baphia massaiensis</i>	1,1	0,7	1,6
<i>Bauhinia macrantha</i>	1,3	1,0	1,3
<i>Burkea africana</i>	0,4	3,1	2,5
<i>Combretum celastroides</i>	0,1	4,5	3,0
<i>C. collinum</i>	0,0	1,2	3,1
<i>C. hereroense</i>	9,3	4,4	1,6
<i>C. zeyheri</i>	0,1	0,7	1,5
<i>Commiphora africana</i>	0,0	2,5	4,5
<i>C. angolensis</i>	0,0	2,2	4,2
<i>C. mozambicensis</i>	0,1	3,1	2,6
<i>C. pyracanthoides</i>	0,0	3,0	4,4
<i>Croton gratissimus</i>	0,1	3,8	3,4
<i>C. pseudopulchellus</i>	0,6	2,8	2,1
<i>Dalbergia melanoxylon</i>	0,0	2,9	4,2
<i>Dichrostachys cinerea</i>	8,5	1,9	1,4
<i>Diospyros lyciodes</i>	0,0	3,8	4,6
<i>D. mespiliformis</i>	0,0	4,0	2,7
<i>Diplorhynchus condylocarpon</i>	0,0	2,1	4,0
<i>Erythrophleum africanum</i>	0,0	2,8	4,3
<i>Grewia flavesces</i>	0,0	0,7	1,7
<i>G. monticola</i>	0,1	1,7	4,0
<i>Guibortia coleosperma</i>	0,0	4,4	1,7
<i>Lonchocarpus nelsii</i>	2,6	1,6	1,9
<i>Ochna pulchra</i>	3,4	1,3	2,2
<i>Pseudolachnostylis maprouneifolia</i>	0,1	4,1	4,4
<i>Rhus tenuinervis</i>	0,0	2,3	3,8
<i>Terminalia sericea</i>	1,3	1,4	2,1
<i>Markhamia acuminata</i>	0,3	4,6	2,2
<i>Pterocarpus angolensis</i>	0,0	3,3	4,7

increase the relative contribution of less abundant species to the total variance. If followed by log. transform, the contributions of these usually under-sampled species becomes so high that they may dominate the analysis.

In summary, the use of unstandardized data, even if zero-transformed, may lead to meaningless abundance gradients. The type of standardization should be such that the rare, less informative species do not play a disproportionately large role in the analysis. Where the range of species values is large a log. transform may be most suitable, as it removes the effects of dominant species but maintains the relative order of importance. The use of a logarithmic transform in plant ecology has been suggested previously by Whittaker (1965). This type of effect can also be achieved by various power transformations, and it may be difficult to determine the optimum transformation procedure.

THE USE OF RELATIVE VALUES

In this procedure the value for a species in a stand is divided by the sum of the values for all species in that stand. The sum of the transformed values in each stand then equals unity. Some authors (Bray and Curtis, 1957; Ayyad and Dix, 1964) believe that relative values offer the best basis for making comparisons between stands, but others (Seal, 1964; Goodall, 1970) believe that the use of absolute values is biologically more meaningful.

Care is initially taken to estimate an importance value which reflects the ecological significance of the species. Values of, say, 100 and 50 for a particular species in two stands are accepted as indicating that the species is ecologically twice as important in the first stand as it is in the second. If relative values are used, the proportions are changed according to the total value of all the other species. Therefore, except where the percentage contribution of each species to a particular stand is of prime importance, absolute values are biologically more meaningful than relative values.

Where environmental measures are included in the analysis, high absolute values for a species in particular stands, corresponding with high values of an environmental factor, may be so reduced by conversion to relative values, as opposed to minor reductions in this species in other stands where there are fewer species, that the correlation between the species and the environmental factor may be obscured.

NON-LINEAR DISTORTIONS

Austin and Noy-meir (1971) and Austin (1972) have discussed the problems of applying linear techniques to data from species with bell-shaped (and therefore non-linear) distributions along environmental gradients, and have demonstrated, with artificial data, the distortions that result. Transform functions exist which can be used to make the data approximate linearity. The difficulty is that in order to know which constants to substitute in the transform equations for each species, one needs some initial idea of the shape of the non-linearity. Generalized non-linear transform functions are needed, and once they have been developed they will no doubt form a basic, initial step for multivariate analyses of most ecological data.

However, untransformed data sets need not necessarily lead to invalid results. In some instances the dominant species in a study do not, in fact, exhibit bell-shaped distributions along major environmental gradients. Two factors can be responsible for this. Firstly, the limits of the study area may exclude one half of the bi-directional curve for many species, although Austin and Noy-Meir (1971) showed that increasing the number of species with such curves decreased distortion only slightly. Secondly, even considering their full range, not all species have bell-shaped distributions. For example, on the

Kalahari sand deposits in Rhodesia the abundance of the dominant tree species (*Baikiaea plurijuga*) increases with increasing sand depth to some as yet undefined level, beyond which it remains at a maximum and does not decrease again to complete the bell-shaped curve. Although admittedly still non-linear, if the number of important species of these two types is large, then the distortions that result will be considerably less than those in the artificial data sets of Swan (1970) and Austin and Noy-Meir (1971).

At this stage in the development of the methods it is important for users of methods that involve ordinations based on Euclidean measures, to be aware of the distortions that can result. Ordination is best suited to the analysis of a comparatively narrow range of vegetation types, and as the range increases so does the amount of distortion due to non-linearity of the species data.

CONCLUSIONS

Consideration of the preceding sections leads to the conclusion that several independent transformations of a set of ecological data are desirable. It is perhaps futile to search for a "best" transformation procedure, when each has particular advantages. Some will result in the analysis being largely influenced by the dominant species and others will emphasize the minor species, each to varying degrees, while still others will place equal importance on all species. More than one transformation may be needed if maximum information is to be gained from a particular procedure.

In relation to the costs of the remainder of the study it is generally worthwhile to perform, or have performed, more than one kind of analysis on more than one transformation of the data, and to compare and combine the information obtained from each.

ACKNOWLEDGEMENTS

Much of this paper was written while I held a Charles Bullard Fellowship for forest research at Harvard University. I thank Drs. J. M. A. Swan and J. White for suggestions and comments on the initial manuscript.

REFERENCES

- AUSTIN, M. P., 1972. Models and analysis of descriptive vegetation data. In: J. N. R. Jeffers (ed.) *Mathematical models in ecology*. British Ecological Society Symposium No. 12 Oxford: Blackwells.
- AUSTIN, M. P. and GREIG-SMITH, P., 1968. The application of quantitative methods to vegetation survey. 1. Some methodological problems of data from rain forest. *J. Ecol.* **56**: 827-844.
- AUSTIN, M. P. and NOY-MEIR, I., 1971. The problem of non-linearity in ordination: experiments with two-gradient models. *J. Ecol.* **59**: 763-773.
- AYYAD, M. A. G. and DIX, R. L., 1964. An analysis of a vegetation-microenvironmental complex on prairie slopes in Saskatchewan. *Ecol. Monogr.* **34**: 421-442.

- BRAY, J. R. and CURTIS, J. T., 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* **27**: 325–349.
- DAGNELIE, P., 1971. Some ideas on the use of multivariate statistical methods in ecology. In: Patil, G. P., Pielou, E. C. and Waters W. E. (eds.) *Statistical Ecology* Vol. 3. Pennsylvania: Penn. State Univ. Press.
- GOFF, F. G. and COTTOM, G., 1967. Gradient analysis: the use of species and sythetic indeces. *Ecology* **48**: 793–806.
- GOODALL, D. W., 1970. Statistical plant ecology. *Annual Review of Ecology and Systematics* **1**: 99–124.
- GREIG-SMITH, P., 1971a. Analysis of vegetation data: the user viewpoint. In: Patil, G. P., Pielou, E. C. and Waters, W. E. (eds.) *Statistical Ecology* Vol. 3. Pennsylvania: Penn. State Univ. Press.
- GREIG-SMITH, P., 1971b. Application of numerical methods to tropical forests. In: Patil, G. P., Pielou, E. C. and Waters, W. E. (eds.) *Statistical Ecology*, Vol. 3. Pennsylvania: Penn. State Univ. Press.
- GRIGAL, D. F. and GOLDSTEIN, R. A., 1971. An integrated ordination—classification analysis of an intensively sampled oak—hickory forest. *J. Ecol.* **59**: 481–492.
- HOPE, K., 1968. *Methods of multivariate analysis*. Univ. London Press.
- LAMBERT, J. M. and DALE, M. B., 1964. The use of statistics in phytosociology. *Advances in Ecological Research*. **2**: 55–99.
- LANGFORD, A. N., and BUELL, M. F., 1969. Integration, identity and stability in the plant association. *Advances in Ecological Research*. **6**: 84–135.
- MOORE, J. J., FITZSIMMONS, S. J. P., LAMBE, E. and WHITE, J., 1970. A comparison and evaluation of some phytosociological techniques. *Vegetatio* **20**: 1–20.
- ORLOCI, L., 1967. Data centering: a review and evaluation with reference to component analysis. *Syst. Zool.* **16**: 208–212.
- PATIL, G. P., PIELOU, E. C. and WATERS, W. E., (eds.) 1971. *Statistical Ecology*, Vol. 1, Vol. 2., Vol. 3. Pennsylvania: Penn. State Univ. Press.
- SEAL, H. L., 1964. *Multivariate Statistical analysis for biologists*. New York: J. Wiley.
- SWAN, J. M. A., 1970. An examination of some ordination problems by use of simulated vegetational data. *Ecology* **51**: 89–102.
- WALKER, B. H. and COUPLAND, R. T., 1970. Herbaceous wetland vegetation in the Aspen Grove and grassland regions of Saskatchewan. *Can. J. Bot.* **40**: 1861–1878.
- WALKER, B. H. and WEHRHAHN, C. F., 1971. Relationships between derived vegetation gradients and measured environmental variables in Saskatchewan wetlands. *Ecology* **52**: 85–95.
- WHITTAKER, R. H., 1965. Dominance and diversity in land plant communities. *Science* **147**: 250–260.
- WILLIAMS, W. T., 1971. Principles of clustering. *Annual Review of Ecology and Systematics*. **2**: 303–326.